Design of RNAi reagents for invertebrate model organisms and human disease vectors

Thomas Horn and Michael Boutros*

German Cancer Research Center (DKFZ), Div. Signaling and Functional Genomics and Heidelberg University, Department of Cell and Molecular Biology, Faculty of Medicine Mannheim, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

*Address for correspondence:
Tel: +49 6221 421951
Fax: +49 6221 421959
Email: m.boutros@dkfz.de

SUMMARY

RNAi has become an important tool to silence gene expression in a variety of organisms, in particular when classical genetic methods are missing. However, application of this method in functional studies has raised new challenges in the design RNAi reagents in order to minimize false positive and false negative results. Since the performance of reagents can be rarely validated on a genome-wide scale, improved computational methods are required that consider experimentally derived design parameters. Here, we describe computational methods for the design of RNAi reagents for invertebrate model organisms and human disease vectors, such as *Anopheles*. We describe procedures how to design short and long double-stranded RNAs for single genes, and evaluate their predicted specificity and efficiency. Using a bioinformatics pipeline we also describe how to design a genome-wide RNAi library for *Anopheles gambiae*.

Keywords: Genomics, RNAi, invertebrates, double-stranded RNA, Drosophila, Anopheles, high-throughput screening

INTRODUCTION

RNA interference (RNAi) screens have become an important tool for the identification and characterization of gene function on a large-scale. They complement classic mutagenesis screens by providing a means to target almost every transcript in a sequenced and annotated genome (1, 2). RNAi is a post-transcriptional gene silencing mechanism conserved from plants to mammals and relies on the delivery of exogenous short double-stranded (ds) RNAs that trigger the degradation of homologous mRNAs in cells (3, 4). As an experimental tool, RNAi is now widely used to silence the expression of genes in a broad spectrum of organisms (5).

RNAi was first observed in plants (6-8) and later mechanistically dissected in the nematode *Caenorhabditis elegans* (4). The introduction of dsRNAs into cells leads to their cleavage into short-interfering RNA (siRNA) duplexes of 21-23 nucleotides length that contain 2-nucleotide 3' overhangs with 5' phosphate and 3' hydroxyl termini (9). This process (referred to as 'dicing') is mediated by the RNase III-like endonuclease Dicer (10-13). The siRNAs are then incorporated into the RNA-induced silencing complex (RISC), and siRNA duplexes are unwound by RISC's helicase activity. The crucial observation that each RISC complex contains only one of the two strands of an siRNA duplex (14), and that only the antisense strand of an siRNA can direct the cleavage of the sense mRNA target has provided important insights into the biochemical mechanism of RNAi-mediated gene silencing (15-20). Thus, it is hypothesized that the RISC preferentially accepts the strand of the siRNA that has the less stable 5' end and that other biophysical, thermodynamic and structural parameters as well as base preferences at specific positions in the sense strand influence the efficiency of incorporation. The siRNA incorporated into the RISC then binds to complementary mRNA that is consequently cut by the RISC RNase H-like nuclease activity at a defined position. Subsequently, the cleaved mRNA is recognized by the cell as aberrant, leading to its degradation (9). This leads to a depletion of the corresponding protein in the cell. The RNAi pathway is not only triggered by exogenous dsRNAs, but also by sequences encoded in the organism's genome, including siRNAs, micro-RNAs (miRNA), and piwi-interacting RNAs. These play important roles in many fundamental biological and disease processes (21).

In *Drosophila*, RNAi can be triggered by 100-700 bp long dsRNAs expressed as hairpins in a time- and tissue-specific manner *in vivo* using the UAS-Gal4 system (22) or added to the culture medium of *Drosophila* cells *in*

vitro (23). The uptake of long dsRNAs in *Drosophila* cells *in vitro* is achieved by receptor-mediated endocytosis (24, 25). In addition, a genome-wide resource to silence transcripts during *Drosophila* oogenesis using short-hairpin RNAs (shRNAs) with the UAS-Gal4 system is being built (26). RNAi-mediated silencing in mammalian cells is mediated through 21-23 nt long siRNAs (27) mimicking Dicer-cleaved products to circumvent an interferon response triggered by long dsRNAs (28). Such short dsRNAs can be generated by different methods. Vectors transcribing short-hairpin RNAs (shRNAs) (29-31) or synthetic siRNAs (27) as well as endoribonuclease-prepared siRNAs (esiRNAs) (32) are commonly used.

The availability of genome-wide RNAi libraries for cell-based assays (5, 33) and whole organisms (34) has opened new avenues to query genomes for a broad spectrum of loss-of-function phenotypes. The number of sequenced invertebrate genomes is steadily rising, enabling reverse genetic approaches using RNAi in many novel model systems, including e.g. the medically relevant vector *Anopheles gambiae* and species used to study evolutionarily aspects of development such as *Tribolium castaneum*, *Acyrthosiphon pisum* and *Schmidtea mediterranea*. RNAi libraries will facilitate the functional characterization of genes in these species, either through studying smaller subsets of candidates or on a genomic scale.

The design of RNAi reagents is key for obtaining reliable phenotypic data in large-scale RNAi experiments. Several recent studies demonstrated that the degradation of non-intended transcripts (so-called ‘off-target effects’) (35, 36) and knock-down efficiency depend on the sequence of the RNAi reagent and have to be carefully monitored (15, 17-20, 37, 38). Based on experimental studies, rules for the design of RNAi reagents have been devised to improve knock-down efficiency and simultaneously minimize unspecific effects.

Here, we provide protocols for the design of RNAi reagents for invertebrate model organisms. We describe a general workflow how to identify suitable target regions that minimize the potential for off-target effects, increase the silencing capacity and allow an efficient synthesis of the reagents. We show how all these steps can be performed automatically, using computational tools that were developed in our group (39, 40). Specifically, we provide case studies for the design of long dsRNAs, siRNAs and shRNAs. Finally, we outline how a genome-wide RNAi resource for a human disease vector genome can be generated.

Materials and Methods

E-RNAi

E-RNAi is a web service available at http://www.e-rnai.org/. It currently enables the design and evaluation of RNAi reagents for 13 organisms, which can be extended upon the request.

NEXT-RNAi

NEXT-RNAi is a software package implemented in Perl and available for download at http://www.nextrnai.org/. It requires the installation of Bowtie (41) and Primer3 (42). To utilize all options of NEXT-RNAi, the BLAST (43), BLAT (44), RNAfold (45) and mdust (*see* Appendix 3) programs are also required. A platform-independent virtual machine (running on VirtualBox) with NEXT-RNAi and all dependencies pre-installed is also available. On a Linux server (two Intel Xeon Quad-Core CPU with 2.00GHz, 16GB RAM) running Ubuntu 9.10 server edition, the design of a genome-wide RNAi library for the *Anopheles* genome took about four hours.

General workflow for the design of RNAi reagents

Selection of suitable RNAi target sequences

1. Obtaining sequences from databases: Potential target sequences (exon, transcript or gene sequences) for RNAi can be downloaded from different genome databases, such as NCBI RefSeq (46) and ENSEMBL (47), or from model-organism databases, such as FlyBase for *Drosophila* (48), BeetleBase for *Tribolium castaneum* (49), Wormbase for *Caenorhabditis elegans* (50) and many others. Table 1 lists invertebrate model organisms amenable to RNAi.
2. Sequences suitable for RNAi synthesis: We recommend using exon sequences as RNAi templates. Sequences containing exons interspersed by small introns also work efficiently. Care needs to be taken for the synthesis of long dsRNAs: if the reagent was designed based on a transcript sequence composed of multiple exons, which are interspersed by long introns in the gene sequences, it should also be amplified from cDNA (not from genomic DNA), since the PCR might fail due to the increased amplicon length. If the reagent was designed based on an exon

sequence, it does not make a difference whether cDNA or genomic DNA is used as a template.

3. Targeting multiple transcripts of the same gene: If the target gene encodes for multiple predicted isoforms, it is recommended to select sequences of exons common to all annotated isoforms, except knocking down a specific isoform is desired.
4. Minimal information about a genome required for RNAi: Novel sequencing technologies allow the rapid assembly of genomes and transcriptomes for emerging model organisms (51, 52). If the organism is amenable to RNAi, this is often the first method of choice for functional studies. How to use such assemblies for the design of RNAi reagents is discussed in details below.

Assessing the predicted RNAi specificity

RNAi reagents have been shown to exert so-called 'off-target' effects, which means they can target additional transcripts besides the desired one. Sequences that are prone to such effects should be excluded from the template sequence, including regions of low complexity, perfect or partial matches to unintended targets and seed region targets:

1. Regions of low complexity: Tandem trinucleotide repeats of the type CA[ACGT] (CAN) are associated with promiscuous off-target effects (84) and longer stretches of such sequences (e.g. more than five contiguous CAN repeats) should be excluded from target regions. Additionally, simple nucleotide repeats and other poly-triplet sequences should also be avoided (see below). Regions of low complexity are particularly critical during the design of long dsRNAs, which are usually longer than 100 bp making them more prone to contain such sequences.
2. Perfect homologies to non-intended transcripts: siRNAs may cause unspecific gene silencing via short stretches of perfect homology with non-intended mRNAs (84-86). To identify such homologies the siRNAs are aligned to the transcriptome of the organism using alignment programs such as BLAST (87) or Bowtie (41) (*see* Appendix 4). For the design of a long dsRNA all possible siRNAs that can be generated need to be aligned. Usually, the long dsRNA is cut into all possible 19-bp long siRNAs (with an offset of 1 base, so that e.g. a 300 bp-long sequence produces 282 siRNAs) to mimic the function of Dicer. A siRNA is predicted to be specific, if it only aligns to transcripts of the intended target gene.
3. Partial homologies to non-intended transcripts: siRNAs can also recognize targets through partial sequence homologies (88, 89). BLAST searches

allowing partial alignments mapping the RNAi reagent to the transcriptome can identify such homologies (see Appendix 5). Significant homologies of a siRNA or a long dsRNA to unintended transcripts should be avoided.

4. miRNA-like seed matches in 3'-UTRs of unintended transcripts: siRNAs can cause unspecific transcript silencing by a route similar to microRNA-mediated silencing through sequence similarity in positions 2-9 (hexamer, heptamer or octamer) of the siRNA guide strand to the 3'-untranslated region (UTR) of non-intended transcripts (90, 91). The number of siRNA seed matches (seed complement frequency) can be determined by aligning the siRNA seed-sequence to a database of 3'-UTR sequences, using e.g. Bowtie. The more targets a seed has, the more likely an off-target silencing effect is (*see* Appendix 6). Thus, siRNA sequences with low seed complement frequencies should be prioritized.

Assessing the predicted RNAi efficiency

Factors influencing the efficiency of RNAi reagents depend very much on the RNAi species. Long dsRNAs, siRNAs and shRNAs are processed in different ways by the cellular RNAi machinery (92). In *Drosophila*, long dsRNAs are first cleaved into 21-23 bp long siRNAs by Dicer2, before they can form a functional RISC. Exogenously applied siRNAs mimic these Dicer2 products and are directly integrated into RISC. shRNA constructs are designed to enter the miRNA biogenesis pathway, including a two-step processing by Drosha/Pasha complexes and Dice 1 into mature shRNAs, before functional RISCs can form.

Every enzymatic step during the processing of exogenous RNAi reagents might influence the overall efficiency. Many studies have been performed in human cell lines to find the best nucleotide composition for siRNAs. Since siRNAs are not processed before they are integrated into RISC, most siRNA properties identified might reflect the preferences of RISC. Several criteria can be taken into account to determine the predicted efficiency of a siRNA:

1. Asymmetric thermodynamic properties: According to the asymmetry rule, the strand (of the double stranded siRNA) with the less stably paired 5' end is incorporated into RISC and becomes the guide strand (17, 19, 93). This should be the strand complementary to the target mRNA, which is called the 'antisense strand'. To increase integration of the antisense strand into RISC, its 5' end should contain A and U bases (in positions 1-5).
2. G/C content: It has been experimentally determined that the G/C content of a functional siRNA should be between 30 and 52% (18, 94). An mRNA

target site with a high G/C content can form stable secondary structures that are not accessible by the siRNA (95).

3. Internal repeats: siRNAs containing internal repeats or palindromes may form internal fold-back structures, which may exist in equilibrium with the duplex form (96) and decrease the effective siRNA concentration and silencing potential. Thus, a functional siRNA should not contain internal repeats that cannot be resolved at temperatures lower or equal to 20°C (18).
4. Base preferences: Different studies identified specific base preferences at several positions required for the optimal efficiency of a siRNA (18, 20, 37). Referring to the sense strand, the following bases are preferred: G/C at position 1, A at position 3, 6 and 19, U at position 10 and 13. The following bases are disadvantaged: A/U at position 1, G at position 13, G/C at position 19.
5. Structural properties of the target: mRNA target sites can take secondary structures with different stabilities making one site more accessible to a siRNA than another (97). Easy accessible target sites should be prioritized.

Two main methods scoring the overall efficiency of a siRNA were developed from these criteria, here referred to as the 'Rational' design by Reynolds et al. (18) and the 'Weighted' design by Shah et al. (38) (*see* Appendix 7). One of the main differences between both methods is that Reynolds et al. suggests the absence of 'internal repeats' a significant feature, whereas Shah et al. does not. Both methods do not consider the structural properties of the target mRNA, which can be calculated using the RNAxs online tool (97).

Recent studies suggest further sequence features influencing RNAi efficiency that could be considered during the reagent design. Argonaut (Ago) proteins form the core of the RISC and possess the endonuclease activity required for target-mRNA cleavage (98). siRNAs bind to a specific Ago family member, Ago2, to mediate RNAi (99). But they can also interact with other Ago proteins (Ago1, Ago3 and Ago4) that usually bind miRNAs. This binding can induce miRNA-like silencing of transcripts by partial homology with their 3'-UTRs, leading to unwanted off-target effects (86, 91, 100-102). It was recently reported that this could be overcome by increasing the siRNA duplex stability, which selectively blocks Ago1, -3 and -4 action and even increases silencing efficiency via Ago2 (103). Duplex stability could be increased by selecting target sites with higher G/C content or by using chemically modified

nucleotides for the passenger strand, to increase thermodynamic stability. However, increasing the G/C content could interfere with target site accessibility. Two other recent studies found siRNA guide strands starting with C preferentially binding to *Drosophila* Ago2, which also increases silencing efficiency and circumvents off-target activities via the other Ago family members (104, 105).

In the case of long dsRNAs it is not understood, which siRNAs are produced by Dicer2. Although there are studies monitoring siRNAs generated from long dsRNAs to study Dicer specificity and efficiency (106), no predictive algorithm is available. However, long dsRNAs have the potential of generating multiple siRNAs, increasing the likelihood that one or more of these can efficiently silence the target transcript. Their efficiency could be predicted using the criteria for siRNAs described above and summarized for the long dsRNA (e.g. by counting the number of efficient siRNAs per long dsRNA or by averaging the efficiency of all siRNAs).

Commonly the same algorithms used to predict the efficiency of siRNAs are used to predict the efficiency of shRNAs (26, 30, 31) and computational tools have been available (107). Using an experimental approach in mammalian cells, it has been shown recently that siRNA algorithms are poor predictors for shRNA efficiency (108). This is most likely because of the multistep process of shRNA biogenesis, which introduces additional structural constraints. For shRNAs, the following efficiency criteria were identified (referring to a 22-bp long sense sequence):

1. An overall high A/U content of 9-18 bases with at least 7 A/U bases between positions 9 and 22.
2. The ratio of percent A/U bases from positions 9-22 and positions 1-8 should be higher 1 to ensure the thermodynamic asymmetry.
3. For antisense strand integration into RISC, there should be an A or U in position 22 of the sense strand, and no U in position 3.
4. There should be either a U or A base at positions 10 or an A base at position 9. At positions 2 and 3 C bases are preferred. This increases Drosha cleavage efficiency.
5. There should be no G in position 3 to increase Drosha cleavage accuracy.
6. An A/U is preferred at position 13 for efficient target cleavage.
7. The sequence should not contain simple nucleotide repeats, such as AAAAAA, UUUUU, CCCC or GGGG.

In addition, the efficiency of shRNA constructs largely depends on the design of the vectors containing the shRNA sequences. These should enable an efficient processing via the miRNA biogenesis pathway (30).

Primer design for long dsRNAs

The amplification of DNA templates for long dsRNAs by PCR from genomic or cDNA sources is a required step during the synthesis of long dsRNAs. Once suitable target sequences with optimal predicted specificity and efficiency have been identified, optimized primers need to be designed.

1. We recommend to use the program Primer3 (42) with default parameters, which is also available as online tool (*see* Appendix 8). The desired 'amplicon size range' (length of dsRNA, e.g. 100-700 bp) should be adjusted, all other parameters might be used with default settings.
2. Primer3 also reports a primer penalty, which scores the overall quality of the primer pair. According to our experience this penalty should be below 10 to enable an efficient PCR synthesis.
3. We have also found that the standardized primer design with Primer3 facilitates similar PCR synthesis efficiency and that the selection of smaller windows for the 'amplicon size range' (resulting in designs of comparable lengths) facilitates similar *in vitro* transcription reactions (*see* Appendix 9).
4. For *in vitro* transcription reactions suitable bacteriophage polymerase promoter sequences, such as T7 (TAATACGACTCACTATAGGG) or SP6 (ATTTAGGTGACACTATAG), must be added to the 5' end of both primers (109).

Ranking of RNAi reagents

The quality of an RNAi target site is determined by the predicted specificity and efficiency as described above. We recommend different strategies for ranking long dsRNAs and siRNAs (for the same target gene) based on the quality measures described.

1. Filtering and ranking of long dsRNAs: Filtering potential target sequences for regions of low sequence complexity should be a first step during the ranking of long dsRNAs. Further, all siRNAs (e.g. 19-bp sequences) contained in a potential RNAi target site should be specific and not show perfect homology to other, non-intended target transcripts. These specific regions can then be used as templates for primer designs with Primer3. Each primer pair will define a potential dsRNA targeting the desired gene. As a last step, the number of predicted efficient siRNAs contained in these dsRNAs should be calculated and the dsRNA with the highest number of

efficient silencers should be prioritized. For some genes it might not be possible to find longer sequence stretches containing only specific siRNAs. In this case, dsRNAs should be first sorted for specificity (prioritizing dsRNAs containing low numbers of unspecific siRNAs) before sorting for efficiency occurs.

2. Filtering and ranking of siRNAs: siRNAs should not contain low-complexity regions and should not show perfect homology to any other than the intended target transcripts. The remaining sequences that pass these filters should be further sorted in the following way: first, for predicted efficiencies (prioritizing efficient silencers) and second for seed complement frequencies (prioritizing siRNAs with small numbers of seed matches to 3'-UTRs of non-intended transcripts).

Case studies

We developed computational tools that enable a straightforward design of optimized long dsRNA-, siRNA- and shRNA-sequences factoring in the quality measures discussed above (Figure 1). E-RNAi (39) is an online tool which facilitates the design of RNAi reagents for currently 13 different organisms. NEXT-RNAi (40) is a stand-alone software that was specifically developed for the design of large-scale RNAi libraries for any sequenced and annotated organism. In the following sections we will provide detailed methods on designing different types of RNAi reagents using E-RNAi. In the last section, we will show how to use NEXT-RNAi for designing novel genome-wide RNAi resources.

Design of long dsRNAs

The design of a long dsRNA needs to take into account both the properties of the target sequence, e,g. its sequence complexity, as well as the properties of all potential siRNAs contained within the long dsRNA, such as their predicted target specificity and efficiency. Because long dsRNAs are often generated by *in vitro* transcription (IVT, see (109)), the design of suitable primer pairs to amplify IVT templates through PCR from genomic DNA or cDNAs must be implemented.

Here, we show how to use E-RNAi (http://www.e-rnai.org/) for the design of long dsRNAs targeting the *Drosophila* gene *flw* (CG2096), which encodes a serine/threonine phosphatase (110).

1. E-RNAi facilitates the selection of target sequences through several ways: using a gene identifier, a gene sequence or by visual selection via the generic genome browser (111). The option 'ID or sequence input' (on the

start page) enables identifier and sequence inputs and provides examples for all available organism. Sequences can be pasted in raw or FASTA formats (*see* Appendix 10). Here, we searched via the identifier *flw*, which is the official gene name. We selected 'long dsRNA' as reagent type and 'D. melanogaster' as organism.

2. E-RNAi identifies the gene model for *flw* as well as all annotated isoforms (FBtr0071446, FBtr0071447) and exons (flw:1-7). From here, target sequences can be selected. We want to target both *flw* isoforms via the exons flw:1 and flw:3-6 (flw:7 equals flw:6 and is not selected) as well as the FBtr0071446-specific isoform flw:2 (Figure 2a).
3. Now E-RNAi allows adjusting several design options (Figure 2c). The default settings E-RNAi uses result from our experience in designing long dsRNAs for *Drosophila* and should be a good starting point for other invertebrates, too.

   Settings (I)-(VII) (Figure 2c) are important to assess the quality of the input sequences. E-RNAi will use these setting as filters:

   I. The siRNA length for specificity prediction is set to 19 bp by default. E-RNAi will align all 19-bp siRNAs derived from the selected exon sequences (with an offset of 1 bp) to the *Drosophila* transcriptome to assess their specificity using Bowtie.
   II. The analysis for regions of low sequence complexity and sequences with more than 5 contiguous CAN repeats are enabled by default. Sequences of low complexity are identified using the mdust program (*see* Appendix 3).
   III. The siRNA seed-match analysis is disabled by default, but allows the user to upload a FASTA file containing 3'-UTR sequences, to adjust the seed length (6, 7 or 8 bases starting at position 2 in the antisense strand) and to define the seed-match cutoff (default 500), which is the maximum allowed number of siRNA seed matches to non-intended 3'-UTRs. The seed complement frequency is then determined for every siRNA calculated from option (I). by aligning the siRNA seeds to uploaded 3'-UTRs using Bowtie. We usually do not use this filter during the design of long dsRNAs, because it will yield hundreds or thousands of predicted 3'-UTR seed matches. It is not predictable, which effect they have on the protein expression of their target.
   IV. The efficiency is predicted for all siRNAs calculated from option (I) and can be calculated using the 'Weighted' or 'Rational' methods (see above). The minimum required efficiency score per siRNA can be adjusted between 0 and 100 to filter for efficient silencers.

According to the 'Weighted' method a score ≥63 defines an efficient siRNA, for the 'Rational' method the score should be ≥66.7 (*see* Appendix 7). We selected the 'Weighted' method and a score of 70 (default is 20). However, it is very unlikely to find a region longer than 100 bp, which only contains efficient siRNAs. To maximize the efficiency, E-RNAi will sort designs according to the number of efficient siRNAs they contain, which is also reported in the output (Figure 3a).

V. The maximum allowed overlap of the long dsRNA with introns (by default 25% of its length in bp) is only important in case genomic sequences containing exons and introns are queried (see point 5 below). It is not important for our example, where exons are selected as target sites.

VI. Settings (I)-(IV) define 'favorable' target regions within the queried *flw* exons. Depending on the stringency applied, these regions might get too short for targeting with a long dsRNA. If 'relaxation of filters' is enabled, E-RNAi can use an iterative approach to generate longer target regions. It will identify the nearest 'favorable' neighbors and include the region in between until the target region is long enough. Consequently, the resulting long dsRNAs might contain unwanted features (e.g. off-target effects). If this option is disabled, the design will fail for these queries.

VII. E-RNAi allows the calculation of optimized long dsRNAs against predicted genes that are not in current annotations (disabled by default). Since *flw* is an annotated gene, this setting is not required here.

Settings (VIII) and (IX) allow further evaluation of long dsRNAs after the design process is finished. Thus, it will not influence their ranking:

VIII. The location of long dsRNAs on the genome will be determined by sequence alignments using BLAT (44) (enabled by default).

IX. The homology of the long dsRNA to sequences in the transcriptome can be determined using BLAST (enabled by default). This setting is different from the specificity prediction in option (I), as it will align the complete sequence to identify partial homologies with non-intended transcripts. The sensitivity can be adjusted by defining the E-value cutoff for the BLAST search (*see* Appendix 5), which is set to 0.00001 by default.

The primer design options define the quality of primer pairs and the desired length of the long dsRNAs. E-RNAi will design primers for all optimal target regions identified from settings (I)-(VII):

X. According to our experience, long dsRNAs as small as 60 bp and as large as 800 bp can be synthesized efficiently and also trigger sufficient knock downs. If multiple long dsRNAs are designed we recommend using the same primer size settings (e.g. the defaults) and a small amplicon size-range (e.g. 200-225 bp). This facilitates comparable PCR and IVT yields as well as transfection efficiencies, when long dsRNAs are transfected during the RNAi experiment (*see* Appendix 9). The primer pair penalty is a measure for the overall quality of the primer pair. Pairs with higher penalty are predicted to perform less efficient during the PCR. According to our experience, penalties up to 10 still allow sufficient PCR yields. The number of primer designs per specific region (default 50) defines the number of primer pairs suggested by Primer3 for each optimal target region. Increasing this value will yield more long dsRNAs covering the target region, also increasing the likelihood to identify better designs, since E-RNAi will finally find the best prediction by sorting them according to their overall quality.

Settings (XI) and (XII) define additional output options:

XI. T7 or SP6 promoters (required for IVT reactions) or any individual tag can be added to the 5'-end of both primers (by default no tag is added).

XII. The number of designs reported per query can be adjusted (default 1). If multiple designs are requested, these do not necessarily target independent regions of the query sequence. Independent designs can be calculated using option (XIII). Further, generic feature files (GFF) and annotation feature files (AFF) for the designed long dsRNA might be requested. These are tab-delimited files that can be used to visualize the location of the long dsRNAs in a genome browser, which is e.g. provided by FlyBase (and other model organism databases).

XIII. Long dsRNAs targeting the *flw* exons designed during this run can be forced to be independent of previously designed reagents targeting *flw*. This requires the upload of FASTA sequences for the available designs. E-RNAi will exclude the regions targeted by these reagents. For this case study no sequences are uploaded.

4. E-RNAi designs long dsRNAs targeting the *flw* exons according to the settings made. A comprehensive report for designs will be generated, which can also be downloaded and locally viewed in any web browser. However, the only design without predicted homologies to non-intended targets is the design targeting exon flw:2, which is specific for transcript FBtr0071446. The main output for this design is shown and described in Figure 3a. Besides the HTML report a tab-delimited file (compatible with any spreadsheet program) containing the same information is also available for download. Sequences are also downloadable as FASTA format. No long dsRNA targeting flw:5 could be calculated, since this exon was too short for the settings made. Long dsRNAs targeting exons flw:1, flw:4 and flw:6 have unwanted 19-bp homologies to non-intended transcripts (Figure 3b), including other members of the PP1 family of phosphatases (*Pp1-87B* and *Pp1α-96A*) and ankyrin repeat-containing proteins (*Ank2*). The design for exon flw:3 has no such perfect 19-bp homologies to unintended transcripts, but shares a significant sequence homology with more than 80% sequence identity over stretches of at least 77 bp with all three members of the PP1 family of phosphatases (*Pp1-87B*, *Pp1α-96A*, *Pp1-13C*) (Figure 3b). This can lead to unspecific knock downs of these phosphatases via partial homology. How an optimal long dsRNA targeting both isoforms of *flw* can be designed by refining the E-RNAi settings is explained next.
5. The exons flw:3-6 are either to short for a long dsRNA or contain short region of homology to non-intended transcripts. On way to target those is to decrease the desired length of the long dsRNA by setting the minimum to 80 bp (for the amplicon size-range in the primer settings), and follow again steps 1-4 described above. However, we sometimes observe decreased synthesis efficiencies for very small amplicons. Since the *flw* exons flw:3-6 are interspersed only by three very small introns, it is possible to include these for the design. Using the GBrowse option of E-RNAi we selected a sequence spanning exons flw:4-6 including the introns in between (Figure 2b). To enable targeting introns, option (V) was modified to allow up to 50% intron content; all other settings were used as described in step 3. This enables E-RNAi to design a specific long dsRNA of suitable size (203 bp) spanning exons flw:4 and flw:5 (Figure 3c). In E-RNAi, introns are subjected to the same quality evaluations as exons, to prevent off-target effects.

Design of siRNAs

Long dsRNAs are the preferred reagents used for RNAi in invertebrate model organisms, because they can be synthesized quite easily (and cheaply) in the lab (109) and show high knock-down efficiencies. However, as the synthesis of siRNAs becomes more cost effective they offer a good alternative for long dsRNA and under certain experimental conditions they even perform more efficient (112). Using siRNAs enables targeting of very short sequences, which might be valuable for transcript-specific knock downs or when very short exons need to be covered. Since siRNA experiments usually use just a single or a pool of three to four siRNAs, the strength of the knock down is prone to non-efficient sequences and the phenotype can be easily affected by targeting of non-intended transcripts. Here, we use E-RNAi to design siRNAs targeting a specific isoform of the *Drosophila* gene *wls/evi* (CG6210), which encodes a transmembrane protein required for Wnt secretion (113).

1. Using FlyBase (48), we downloaded the sequence for *wls* exon 2 (wls:2, 96 bp long), an exon specific to the transcript FBtr0076218, which is one of the two annotated transcripts of *wls*. The sequence was pasted in FASTA format in the sequence box using the 'ID or sequence input option', 'siRNA' was selected as reagent type and 'D. melanogaster' as organism.
2. E-RNAi aligns the input sequence to the *Drosophila* genome using BLAT and displays the mapping in GBrowse (Figure 4a).
3. Now E-RNAi allows adjusting the design options (see Figure 2c for long dsRNA designs). All settings described for the design of long dsRNAs in section 3.2.1 also apply for the design of siRNAs (except for the options regarding primer design and tagging), but have different default settings. Importantly, the minimum efficiency in option (IV) is set to a score of 70 using the 'Weighted' method. siRNAs with this score are predicted to be efficient silencers. Further, the default E-value cutoff for evaluating the homology of designed siRNAs in option (IV) is 0.1, following the smaller size of siRNAs. One option, that was not used during the design of long dsRNAs is the siRNA seed match analysis, which is more important during the design of siRNAs:

   III. We downloaded a pre-build file containing 18,807 annotated 3'-UTRs from FlyBase. This file was uploaded to E-RNAi as FASTA database to identify siRNA seed matches. The seed match cutoff was set to 500 for aligning octamers (positions 2-9 in siRNA antisense strand). It has been shown that siRNAs with low seed complement frequencies induce fewer off-target

phenotypes than siRNAs with more abundant 3'-UTR complements (90) (*see* Appendix 6).

VII. We requested four siRNAs to be designed. They will be independent and could be used in a pool for an experiment.

4. E-RNAi designs four siRNAs targeting exon 2 of *wls*. The output is structured in a similar way as for long dsRNAs and exemplified for one siRNA in Figure 4b. All four siRNAs are specific for there intended target, have an efficiency score of at least 75 and three out of four siRNAs have seed complement frequencies of less than 300.
5. E-RNAi reports the 'raw' siRNA sense sequence. For RNA synthesis, 'Ts' would need to be changed to 'Us' and 2-nucleotide overhangs (usually 'TT') would need to be added to the 3'-end of both, the sense and antisense strands. Companies further developed chemical modifications that are supposed to increase the efficiency of the antisense strand or to avoid off-target effects arising from the sense strand.

Design of shRNAs

Recently, a novel transgenic RNAi resource based on shRNAs has been generated for *Drosophila* (26) using a design method optimized for siRNAs (107) to predict efficient sequences. In this study, the shRNAs tested did not only proof to be very effective at silencing gene expression, but they also could be used to knock down transcripts in the female germline, where resources based on long dsRNAs do not function. shRNA libraries have a longer history in mammalian systems, because compared to siRNAs they can be delivered into any cell type using lentiviral transfection and greatly facilitate pooled screening approaches (30, 31). However, recent studies found that shRNAs have sequence requirements different from siRNAs making siRNA efficiency rules poor predictors for shRNA efficiency (108, 114, 115).

E-RNAi in its current version does not implement an shRNA-specific efficiency predictor. The sequence preferences derived from a current study (108) are described above and could be used for identifying efficient sequences. Sequences that pass the efficiency filter could then be evaluated for their specificity using the 'Evaluation' option of E-RNAi. This requires just pasting the sequences in FASTA format, selecting 'siRNA' as reagent type and the desired organisms. Similarly to the design of RNAi reagents, E-RNAi offers different options for sequence evaluations. However, default settings will work fine. Only the length of the siRNA/shRNA for specificity evaluation needs to be adjusted properly (19 bp by default).

Depending on the application, shRNAs are usually cloned into specific vectors. Consequently, any restriction sites required for cloning must be excluded from the shRNA sequence.

Design of independent RNAi reagents

In order to validate phenotypes from RNAi experiments, validation experiments using independent RNAi reagents are recommended (35). E-RNAi provides different ways to design independent RNAi reagents:

1. The sequence of a previously designed reagent can be submitted as FASTA file in the process of designing a new reagent (see section 3.2.1, option (XIII)). E-RNAi will exclude this sequence during the design.
2. E-RNAi can be queried with multiple sequences for one run, so e.g. multiple exons of the same transcript can be queried and the resulting designs will be independent. Similarly, after querying E-RNAi with a sequence identifier, all annotated exon sequences for this gene are shown. Multiple sequences can be selected resulting in independent designs. For the design of independent siRNAs, selecting more than one design for output in design option (VII) will be sufficient to obtain independent reagents.
3. The possibility to submit sequences from GBrowse to E-RNAi allows a precise selection of potential target sequences (see Figure 2b). Multiple, independent sequences can be submitted this way resulting in independent designs.

Design of genome-wide RNAi libraries

We developed the software package NEXT-RNAi that allows the design and evaluation of genome-wide RNAi libraries using the workflow shown in Figure 1 (40). NEXT-RNAi is a standalone Perl program and depends on several additional programs (see Materials). Here, we describe how NEXT-RNAi can be used to design a novel library of long dsRNAs for the mosquito *A. gambiae*, which is widely studied to analyze the mechanism of innate immunity as a vector for *Plasmodium falciparum*. RNAi by long dsRNAs has been demonstrated *in vitro* and *in vivo* and leads to efficient depletion of mRNAs (69).

1. Obtaining sequence and sequence-feature files: VectorBase is the model organism database for invertebrate vectors of human pathogens, including *A. gambiae* (68). NEXT-RNAi requires a defined set of input files with fixed formats. Those can either be downloaded

directly from VectorBase or generated via its database management system BioMart (116), a tool that is becoming available for an increasing number of model organism databases to facilitate sequence and feature downloads. The following files are required:

I. File containing target sequences: we recommend using exon sequences as target sites. This facilitates amplification of templates for long dsRNAs from genomic DNA and cDNA. If whole transcripts are used as target sites, cDNA libraries are required for the initial amplification step. Exons sequences in FASTA format can be downloaded via the BioMart implementation of VectorBase. We identified ~56,000 exon sequences with an average length of 418 bp, offering suitable target sites. To obtain even more potential target sequences, we splitted all exons longer than 560 bp into two halves, resulting in ~76,600 sequences.

II. Off-target transcript database: In order to assess the specificity of target sequences, NEXT-RNAi requires a database containing all annotated transcripts. This file (in FASTA format) can be directly downloaded from VectorBase. It contains ~15,000 transcript sequences. NEXT-RNAi uses Bowtie for aligning siRNAs to the transcriptome. Bowtie requires building of an index from the transcriptome FASTA file. This can be done using the *bowtie_build* script included in the Bowtie software. The Bowtie webpage also offers many pre-build indices, which can be downloaded and directly used.

    In order to evaluate the designed reagents for partial homologies to unintended transcripts a BLAST index needs to be build from the transcriptome FASTA file using the *formatdb* script included in the BLAST program.

III. Linking transcripts and genes: NEXT-RNAi needs to know, which annotated transcripts belong to the same gene in order to discriminate between on- and off-target effects. This requires a simple tab-delimited file with two columns: the column ‘Target’ contains the transcript identifier and the column ‘TargetGroup’ contains the corresponding gene identifier (for *Anopheles* e.g. the transcript ‘AGAP004677-RB’ belongs to the gene AGAP004677). This file can be generated using BioMart, querying for ‘Ensembl Transcript ID’ and ‘Ensembl Gene ID’.

Only the header names need to be changes to ‘Target’ and ‘TargetGroup’ respectively.

These are the minimal input requirements for NEXT-RNAi. In order to align reagents to the genome, a FASTA file containing the *Anopheles* chromosomes would be required (and build into a Bowtie index), which is also available from VectorBase. Additional sequence features could also be analyzed, e.g. for whether the sequences contain annotated SNPs or conserved miRNAs seeds. Examples, how to prepare such files are available from the NEXT-RNAi wiki (*see* Appendix 11).

2. Adjusting the design options: The user can modify basically every step during the design of RNAi reagents implemented in NEXT-RNAi. Although most of these can be used with default values, we recommend adjusting the following options:
   I. The desired minimum and maximum length of the long dsRNA (we recommend to use a rather small window, e.g. 200-225 bp).
   II. Enable the analysis for low-complexity regions (simple nucleotide repeats and ≥6x CAN repeats).
   III. The siRNA length for the off-target analysis (we recommend 19 bp).
   IV. Definition of the efficiency algorithm (‘Weighted’ or ‘Rational’) used for assessing the siRNA efficiency and the efficiency cutoff. For long dsRNAs, we recommend to set the cutoff to 0. This will not set any restrictions to the efficiency of individual siRNAs, but NEXT-RNAi will finally sort all potential designs according to their average efficiency.
   V. Enable the redesign option, which allows NEXT-RNAi to relax specificity and efficiency parameters.
   VI. Enable the analysis for homology to non-intended transcripts.
   VII. Adjust the number of long dsRNAs to be designed per input sequence (in this case per exon). We recommend just designing one sequence per exon.
   VIII. Provide the location of the file linking transcripts and genes.

These options can be set using an options file (with a *OPTION=VALUE* structure) and provided NEXT-RNAi during its start. Alternatively, NEXT-RNAi offers and interactive starting mode, which helps to defined these settings. Further options and examples are explained in the NEXT-RNAi wiki (*see* Appendix 11).

3. Starting the design: NEXT-RNAi is a Perl script that can be started e.g. from command-line. It requires some additional information for the startup, including (i) the location of the sequence input-file and (ii) the information, whether these sequences are templates for a *de novo* design or the evaluation of RNAi reagents, (iii) the type of reagents to be designed, (iv) the location of the Bowtie index of the transcriptome, (v) the location of the options file, and (vi) an identifier for the design process (e.g. 'agambiae'). Help about how to start NEXT-RNAi and all available options can be obtained starting NEXT-RNAi with '-h'.
4. The output: NEXT-RNAi produces a comprehensive HTML output providing links to detailed reports for every long dsRNA designed (see e.g. Figure 3 and Appendix 12). Further, a tab-delimited file (compatible with every spreadsheet software) containing the same information for all designs is provided.
5. Working with the output: Overall, NEXT-RNAi designed 48,128 long dsRNAs covering 12,286 of the 13,319 annotated genes (92%). About 89% of the targeted genes are covered by a long dsRNA with no predicted off-target effects and 86% are even covered by at least two independent reagents. We recommend identifying the best two independent designs per gene by sorting the long dsRNAs for their target gene, followed by sorting steps for specificity and efficiency. In addition, reagents that target other, non-intended genes (either via 19-bp homologies or significant overall homology) or that contain low-complexity regions should be filtered out. These steps could be easily accomplished using spreadsheet software. Next, to reach a nearly complete coverage, genes that were not covered so far or are just covered by a long dsRNA with predicted off-target effects as well as genes that are only covered by one long dsRNA could be repeated in a second run. For this run we recommend to decrease the minimum amplicon length allowed (e.g. to 80 bp). This will allow NEXT-RNAi to target smaller regions. Further, for genes were a second, independent design is desired, the first designs should be provided to NEXT-RNAi. These sequences will be excluded for targeting by a second reagent. After this run, a library with nearly complete genome coverage by two independent designs per gene will be achieved.
6. Synthesizing the library: The long dsRNAs designed for the *Anopheles* genome could be synthesized by amplification of DNA templates from genomic DNA or cDNA using PCRs, followed by *in vitro* transcription of the DNA using e.g. T7 polymerase. Primers designed by NEXT-RNAi

all have similar properties (e.g. melting temperatures and GC content) facilitating efficient synthesis in high-throughput formats, such as 96-well plates. To this end, we recommend using a two-step PCR approach for DNA amplification, which requires tagging of primers with re-amplifiable tags (109).

Appendix

1. If genes have multiple annotated isoforms, regions common to all of them are very likely to be real. We identify such regions by comparing the absolute chromosomal locations of all exons (and UTRs) of all transcripts of a certain gene. Parts of the exons that exist in all isoforms are considered potential target sites. Common regions that are too short for the design of a long dsRNA (e.g. <120 bp) can also be extended on both ends (e.g. by 70 bp). Very long common regions (e.g. >700 bp) can be split into two halves to provide more independent target sites. Targeting specific isoforms of a gene requires the identification of exons or UTRs that are unique for this transcript.
2. The optimal design of RNAi reagents requires annotations for the genome and the transcriptome, including the information about which transcripts belong to the same gene. Gene functions can then be annotated by BLAST searches to other, well-annotated organisms, preferentially related species. This allows the identification of target genes with desired functions. Then, based on the transcript sequences, RNAi reagents can be designed. However, in the case of long dsRNAs, which requires synthesis by PCRs, DNA templates would need to be amplified from cDNA (not from genomic DNA), since regions in the transcripts might be interspersed by large introns. If assemblies of the genome (contigs or even chromosomes) are available this could be tested by aligning the long dsRNAs using BLAT.
3. We analyze sequences for low-complexity regions using the so-called 'mdust' filter. This filter is available as BioPerl Module (at http://www.bioperl.org/wiki/Mdust) and as standalone C script (at http://compbio.dfci.harvard.edu/tgi/software/ or via our NEXT-RNAi packed archives at http://www.nextrnai.org/). We use the software with default settings.
4. For aligning many siRNA sequences to the transcriptome or genome we recommend to use Bowtie (http://bowtie-bio.sourceforge.net/), which was developed for aligning short sequences from next-generation sequencing data. We use it with the options '-f' (reading from FASTA files, *see* Appendix 10), '-v 0' (allowing no mismatches). BLAST could also be used, however, Bowtie is significantly faster for aligning short sequences.
5. To determine the overall homology of an RNAi reagent to all annotated transcripts we use BLAST to align the sequence to the transcriptome.

This will uncover partial homologies to non-intended transcripts. Except excluding low-complexity sequences ('-F F') and setting an E-value cutoff (long dsRNAs: 0.00001, siRNAs: 0.1) we use default BLAST parameters. The low E-value cutoffs we use avoid reporting of non-significant homologies. However, other tools such as WU-BLAST (now AB-BLAST) might be even more sensitive for alignments of very short sequences considering mismatches.

6. We only consider seed complement frequencies during the design of siRNAs, since analyzing the seed matches for all siRNAs contained in a long dsRNA will be very time consuming and result in an immense amount of potential off-target effects. But even for single siRNAs the predictive value of the seed complement frequency is rather weak. Anderson et al. (90) found a correlation of low seed complement frequencies (<350) with a small number of off-target effects and medium seed complement frequencies (~2500) with a high number of off-target effects. However, for high seed complement frequencies (>3800) the number of off-target effects decreased again. Further, even if seed complement frequencies are low, it is not clear, which of the targeted transcripts are really silenced.
7. The 'Rational' prediction includes eight criteria to score the efficiency of a siRNA (sense strand): (i) G/C content (+1 if between 30% and 52%), (ii) A/U bases at positions 15-19 (+1 each), (iii) absence of internal repeats (+1 if melting temperature of potential internal hairpin < 20°C), (iv) A base at position 19 (+1), (v) A base at position 3 (+1), (vi) U base at position 10 (+1), (vii) G/C base at position 19 (-1) and (viii) G base at position 13 (-1). This scoring provides a score range from -2 to 10. In the studies done by Reynolds et al. (18) siRNAs with scores ≥6 were found to be efficient silencers. We normalized this score to obtain a range between 0 and 100 using the following formula:

   FinalScore = ((Score of siRNA - minimal possible score) / (maximal possible score - minimal possible score)) * 100

   For a siRNAs considered an efficient silencer its score must be higher 66.67.

   The 'weighted' prediction includes twelve criteria to predict the efficiency of a siRNA (sense strand): (i) A/U base at position 1 (-1.4), (ii) G/C base at position 1 (+1.11), (iii) A base at position 6 (+0.7), (iv) U base at position 10 (+0.25), (v) G base at position 13 (-1.66), (vi) U base at position 13 (+0.31), (vii-x) A/U bases at position 16-19 (+0.74, +1.2, +1.44, +0.87), (xi) G/C base at position 19 (-1.02) and (xii) G/C

content (+0.42 if between 30% and 55%). This scoring provides a score range from -4.08 and 7.04. We applied the same normalization to obtain a score range between 0 and 100 as described for the 'Rational' score. Shah et al. (38) found in their study that the average score of the most potent siRNAs was 63.

8. The Primer3 online tool is available at http://frodo.wi.mit.edu/primer3/.
9. We currently design long dsRNAs in the size range of 200 and 225 bp if possible (and decrease the minimum allowed size otherwise). We found that this results in more similar PCR and IVT yields. In case transfection reagents (e.g. lipid-based transfection) are used for RNAi delivery into the cells, this also facilitates similar transfection efficiencies.
10. The FASTA format is a general sequence format composed of a header starting with '>' followed by the sequence identifier. The next line contains the sequence. More detailed information is available at http://en.wikipedia.org/wiki/FASTA_format. 'Raw' sequence means that the sequence is provided without any identifier.
11. Detailed descriptions how to use the E-RNAi online service (http://www.e-rnai.org) are available via the wiki at http://b110-wiki.dkfz.de/signaling/wiki/display/ernai/. The NEXT-RNAi software, installation packages and instructions for Linux and Mac operation systems and further documentations are accessible via http://www.nextrnai.org/.
12. A detailed description for the preparation of files required for designing a genome-wide library of long dsRNAs for *Anopheles* and instructions how to run NEXT-RNAi with these files is available at http://b110-wiki.dkfz.de/signaling/wiki/display/nextrnai/Anopheles+gambiae.

Figure legends

Figure 1. Computational workflow of RNAi design.
Target sequence files and sequence databases are provided in FASTA format. The target sequences are fist filtered for six (default) or more contiguous CAN repeats and for other regions of low complexity (e.g., simple nucleotide repeats) using mdust (*see* **Appendix 3**). Sequences are then 'diced' to generate all possible siRNA sequences with a default length of 19 bp and an offset of 1 bp. Subsequently, each siRNA is aligned to a user-defined 'off-target' database (e.g., the whole transcriptome) to determine its specificity using Bowtie. The specificity is set to 1 if the siRNA targets a single gene or to 0 otherwise. In the next step, the predicted efficiency of each 19 bp siRNA is computed using either the 'Rational' method according to Reynolds et al. (18) or the 'weighted' method according to Shah et al. (38), assigning each siRNA an efficiency score between 0 and 100. Optionally, the seed complement frequency for each siRNA can be computed for any FASTA file provided (e.g., a file containing 3'-UTR sequences). siRNAs that did not pass the low-complexity filters, show perfect homology to multiple target genes or do not meet the user-defined cutoffs for efficiency or seed complement frequency are excluded from the queried target sequences. Remaining sequences are used as templates for primer design with Primer3 for long dsRNAs or are directly subjected to the final ranking for the design of siRNAs. Designs are ranked by (i) their predicted specificity and (ii) their predicted efficiency and, in the case of siRNA designs, (iii) their calculated seed complement frequency (preferring a low number of seed matches). Sequences can also be evaluated for additional features, such as homology to non-intended transcripts, or SNP- and UTR-contents. Final designs can be visualized using GBrowse. All results are presented in a comprehensive HTML report and are also exported to text files.

Figure 2. Sequence selection and design options.
(a) After querying E-RNAi with gene identifiers, gene sequences, transcript sequences or exon sequences can be selected for design. (b) Individual sequences can be selected as design templates visually from GBrowse (upper panel). After sequence submission to E-RNAi, the type of reagent (long dsRNA or siRNA) can be selected (lower panel). (c) Options available for the design of long dsRNAs include settings for the prediction of specificity, efficiency and low-complexity regions (left panel) as well as primer design settings (upper right panel). In addition different output options, such as the

number of designs per query sequence and different report formats, can be adjusted. The sequence of a previous design can also be uploaded in FASTA format to exclude it from a new (independent) design (lower right panel). Details for certain options are provided in the text.

Figure 3. Report for the design of a long dsRNA against the gene *flw*.
(a) Optimal design for exon flw:2: 'dsRNA information' summarizes the properties of the primers required to amplify the reagent from genomic or cDNA sources and also shows the full sequence of the construct, its length and location in the genome. The primers shown here were also tagged with the T7 promoter sequence (lower letters) in 5'. 'Target information' lists all 'intended' and 'other' target genes found as well as all transcripts belonging to them (including the number of siRNA 'hits' to each transcript). 'Reagent quality' summarizes the analyzed quality parameters and shows the number of specific ('on-target') siRNAs contained within the dsRNA, the number of 'off-target' and 'no-target' siRNAs as well as the number of 'efficient siRNAs', the average siRNA efficiency and low-complexity information. In addition, all targets with significant 'sequence homology' to the long dsRNA are listed. GBrowse visualizes the location of the designed reagent in the gene model. (b) Suboptimal designs for exons flw:1, -3, -4, -6: for these designs only parts of the reports are shown. Designs for exons flw:1, -4 and -6 all have predicted 19-bp homologies to unintended transcripts (red boxes), which can be read from the 'Target information' and 'Reagent quality' boxes. The design for flw:3 shows significant overall homology to unintended transcripts (red box), which can be read from the 'Additional quality evaluation' box. (c) Optimal design for the region spanning *flw* exons 3-6: the design spans the intron between exon 4 and 5 (74 19-bp siRNAs have 'No target') and has no predicted off-target effects.

Figure 4. Report for the design of siRNAs against the gene *wls*.
(a) Selection of *wls* exon 2 as target site, which is specific for the transcript FBtr0076218. (b) Optimal siRNA design for the *wls* exon: 'siRNA information' shows the 19-bp long siRNA sequence, its position in the input sequence as well as its position in the genome. 'Target information' lists all 'intended' and 'other' target genes found. This siRNA is specific for *wls* and targets the transcript FBtr0076218, but not FBtr0076219. 'Reagent quality' summarizes the analyzed quality parameters and shows that this siRNA is specific ('on-target'), has 150 seed matches ('SCF', seed complement frequency) in annotated 3'-UTRs, an efficiency score of ~82 and contains no low-complexity

sequences (‘LowComplexRegions’ or ‘CAN’ repeats). Further, there is no ‘sequence homology’ to unintended genes. The genome browser image visualizes its location.

References

1. Jorgensen, E. M., and Mango, S. E. (2002) The art and design of genetic screens: caenorhabditis elegans, *Nat Rev Genet **3***, 356-369.
2. St Johnston, D. (2002) The art and design of genetic screens: Drosophila melanogaster, *Nat Rev Genet **3***, 176-188.
3. Chapman, E. J., and Carrington, J. C. (2007) Specialization and evolution of endogenous small RNA pathways, *Nat Rev Genet **8***, 884-896.
4. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature **391***, 806-811.
5. Boutros, M., and Ahringer, J. (2008) The art and design of genetic screens: RNA interference, *Nat Rev Genet **9***, 554-566.
6. van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N., and Stuitje, A. R. (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression, *The Plant cell **2***, 291-299.
7. van der Krol, A. R., Mur, L. A., de Lange, P., Mol, J. N., and Stuitje, A. R. (1990) Inhibition of flower pigmentation by antisense CHS genes: promoter and minimal sequence requirements for the antisense effect, *Plant molecular biology **14***, 457-466.
8. Napoli, C., Lemieux, C., and Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans, *The Plant cell **2***, 279-289.
9. Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, *Cell **101***, 25-33.
10. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference, *Nature **409***, 363-366.
11. Bernstein, S. L., Guo, Y., Kelman, S. E., Flower, R. W., and Johnson, M. A. (2003) Functional and cellular responses in a novel rodent model of anterior ischemic optic neuropathy, *Investigative ophthalmology & visual science **44***, 4153-4162.
12. Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells, *Nature **404***, 293-296.
13. Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans, *Genes & development **15***, 2654-2659.
14. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi, *Cell **110***, 563-574.
15. Chiu, Y. L., and Rana, T. M. (2002) RNAi in human cells: basic structural and functional features of small interfering RNA, *Molecular cell **10***, 549-561.

16. Chiu, Y. L., and Rana, T. M. (2003) siRNA function in RNAi: a chemical modification analysis, *RNA (New York, N.Y* ***9***, 1034-1048.
17. Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias, *Cell* ***115***, 209-216.
18. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., and Khvorova, A. (2004) Rational siRNA design for RNA interference, *Nature biotechnology* ***22***, 326-330.
19. Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003) Asymmetry in the assembly of the RNAi enzyme complex, *Cell* ***115***, 199-208.
20. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R., and Saigo, K. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference, *Nucleic acids research* ***32***, 936-948.
21. Liu, Q., and Paroo, Z. (2010) Biochemical principles of small RNA pathways, *Annu Rev Biochem* ***79***, 295-319.
22. Fischer, J. A., Giniger, E., Maniatis, T., and Ptashne, M. (1988) GAL4 activates transcription in Drosophila, *Nature* ***332***, 853-856.
23. Clemens, J. C., Worby, C. A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B. A., and Dixon, J. E. (2000) Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways, *Proceedings of the National Academy of Sciences of the United States of America* ***97***, 6499-6503.
24. Ulvila, J., Parikka, M., Kleino, A., Sormunen, R., Ezekowitz, R. A., Kocks, C., and Ramet, M. (2006) Double-stranded RNA is internalized by scavenger receptor-mediated endocytosis in Drosophila S2 cells, *J Biol Chem* ***281***, 14370-14375.
25. Saleh, M. C., van Rij, R. P., Hekele, A., Gillis, A., Foley, E., O'Farrell, P. H., and Andino, R. (2006) The endocytic pathway mediates cell entry of dsRNA to induce RNAi silencing, *Nature cell biology* ***8***, 793-802.
26. Ni, J. Q., Zhou, R., Czech, B., Liu, L. P., Holderbaum, L., Yang-Zhou, D., Shim, H. S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L. A., Hannon, G. J., and Perrimon, N. (2011) A genome-scale shRNA resource for transgenic RNAi in Drosophila, *Nature methods* ***8***, 405-407.
27. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* ***411***, 494-498.
28. Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H., and Williams, B. R. (2003) Activation of the interferon system by short-interfering RNAs, *Nature cell biology* ***5***, 834-839.
29. Bernards, R., Brummelkamp, T. R., and Beijersbergen, R. L. (2006) shRNA libraries and their use in cancer genetics, *Nature methods* ***3***, 701-706.
30. Chang, K., Elledge, S. J., and Hannon, G. J. (2006) Lessons from Nature: microRNA-based shRNA libraries, *Nature methods* ***3***, 707-714.
31. Root, D. E., Hacohen, N., Hahn, W. C., Lander, E. S., and Sabatini, D. M. (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library, *Nature methods* ***3***, 715-719.

32. Buchholz, F., Kittler, R., Slabicki, M., and Theis, M. (2006) Enzymatically prepared RNAi libraries, *Nature methods* ***3***, 696-700.
33. Fuchs, F., and Boutros, M. (2006) Cellular phenotyping by RNAi, *Briefings in functional genomics & proteomics* ***5***, 52-56.
34. Dietzl, G., Chen, D., Schnorrer, F., Su, K. C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., and Dickson, B. J. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila, *Nature* ***448***, 151-156.
35. Echeverri, C. J., Beachy, P. A., Baum, B., Boutros, M., Buchholz, F., Chanda, S. K., Downward, J., Ellenberg, J., Fraser, A. G., Hacohen, N., Hahn, W. C., Jackson, A. L., Kiger, A., Linsley, P. S., Lum, L., Ma, Y., Mathey-Prevot, B., Root, D. E., Sabatini, D. M., Taipale, J., Perrimon, N., and Bernards, R. (2006) Minimizing the risk of reporting false positives in large-scale RNAi screens, *Nature methods* ***3***, 777-779.
36. Perrimon, N., and Mathey-Prevot, B. (2007) Off Targets and Genome Scale RNAi Screens in Drosophila, *Fly* ***1***, 5.
37. Amarzguioui, M., and Prydz, H. (2004) An algorithm for selection of functional siRNA sequences, *Biochemical and biophysical research communications* ***316***, 1050-1058.
38. Shah, J. K., Garner, H. R., White, M. A., Shames, D. S., and Minna, J. D. (2007) sIR: siRNA Information Resource, a web-based tool for siRNA sequence design and analysis and an open access siRNA database, *BMC Bioinformatics* ***8***, 178.
39. Horn, T., and Boutros, M. (2010) E-RNAi: a web application for the multi-species design of RNAi reagents--2010 update, *Nucleic acids research* ***38***, W332-339.
40. Horn, T., Sandmann, T., and Boutros, M. (2010) Design and evaluation of genome-wide libraries for RNA interference screens, *Genome Biol* ***11***, R61.
41. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol* ***10***, R25.
42. Rozen, S., and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers, *Methods Mol Biol* ***132***, 365-386.
43. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J Mol Biol* ***215***, 403-410.
44. Kent, W. J. (2002) BLAT--the BLAST-like alignment tool, *Genome research* ***12***, 656-664.
45. Hofacker, I. L. (2003) Vienna RNA secondary structure server, *Nucleic acids research* ***31***, 3429-3431.
46. Pruitt, K. D., Tatusova, T., Klimke, W., and Maglott, D. R. (2009) NCBI Reference Sequences: current status, policy and new initiatives, *Nucleic acids research* ***37***, D32-36.
47. Flicek, P., Aken, B. L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Koscielny, G., Kulesha, E., Lawson, D.,

Longden, I., Massingham, T., McLaren, W., Megy, K., Overduin, B., Pritchard, B., Rios, D., Ruffier, M., Schuster, M., Slater, G., Smedley, D., Spudich, G., Tang, Y. A., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S. P., Zadissa, A., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Smith, J., and Searle, S. M. (2010) Ensembl's 10th year, *Nucleic acids research* ***38***, D557-562.

48. Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., and Zhang, H. (2009) FlyBase: enhancing Drosophila Gene Ontology annotations, *Nucleic acids research* ***37***, D555-559.
49. Kim, H. S., Murphy, T., Xia, J., Caragea, D., Park, Y., Beeman, R. W., Lorenzen, M. D., Butcher, S., Manak, J. R., and Brown, S. J. (2010) BeetleBase in 2010: revisions to provide comprehensive genomic information for Tribolium castaneum, *Nucleic acids research* ***38***, D437-442.
50. Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H. M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L. D., Spieth, J., and Sternberg, P. W. (2010) WormBase: a comprehensive resource for nematode research, *Nucleic acids research* ***38***, D463-467.
51. Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics, *Nature reviews. Genetics* ***10***, 57-63.
52. Metzker, M. L. (2010) Sequencing technologies - the next generation, *Nature reviews. Genetics* ***11***, 31-46.
53. Legeai, F., Shigenobu, S., Gauthier, J. P., Colbourne, J., Rispe, C., Collin, O., Richards, S., Wilson, A. C., Murphy, T., and Tagu, D. (2010) AphidBase: a centralized bioinformatic resource for annotation of the pea aphid genome, *Insect Mol Biol* ***19 Suppl 2***, 5-12.
54. Mutti, N. S., Park, Y., Reese, J. C., and Reeck, G. R. (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum, *J Insect Sci* ***6***, 1-7.
55. Jaubert-Possamai, S., Le Trionnaire, G., Bonhomme, J., Christophides, G. K., Rispe, C., and Tagu, D. (2007) Gene knockdown by RNAi in the pea aphid Acyrthosiphon pisum, *BMC Biotechnol* ***7***, 63.
56. Munoz-Torres, M. C., Reese, J. T., Childers, C. P., Bennett, A. K., Sundaram, J. P., Childs, K. L., Anzola, J. M., Milshina, N., and Elsik, C. G. (2011) Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera, *Nucleic acids research* ***39***, D658-662.
57. Jarosch, A., and Moritz, R. F. (2011) Systemic RNA-interference in the honeybee Apis mellifera: tissue dependent uptake of fluorescent siRNA after intra-abdominal application observed by laser-scanning microscopy, *J Insect Physiol* ***57***, 851-857.

58. Dearden, P. K., Duncan, E. J., and Wilson, M. J. (2009) RNA interference (RNAi) in honeybee (Apis mellifera) embryos, *Cold Spring Harb Protoc* ***2009***, pdb prot5228.
59. Lee, J. A., Kim, H. K., Kim, K. H., Han, J. H., Lee, Y. S., Lim, C. S., Chang, D. J., Kubo, T., and Kaang, B. K. (2001) Overexpression of and RNA interference with the CCAAT enhancer-binding protein on long-term facilitation of Aplysia sensory to motor synapses, *Learn Mem* ***8***, 220-226.
60. Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z., and Xia, Q. (2010) SilkDB v2.0: a platform for silkworm (Bombyx mori) genome biology, *Nucleic acids research* ***38***, D453-456.
61. Masumoto, M., Yaginuma, T., and Niimi, T. (2009) Functional analysis of Ultrabithorax in the silkworm, Bombyx mori, using RNAi, *Dev Genes Evol* ***219***, 437-444.
62. Song, C., Gallup, J. M., Day, T. A., Bartholomay, L. C., and Kimber, M. J. (2010) Development of an in vivo RNAi protocol to investigate gene function in the filarial nematode, Brugia malayi, *PLoS Pathog* ***6***, e1001239.
63. Maine, E. M. (2008) Studying gene function in Caenorhabditis elegans using RNA-mediated interference, *Briefings in functional genomics & proteomics* ***7***, 184-194.
64. Nishiyama, A., and Fujiwara, S. (2008) RNA interference by expressing short hairpin RNA in the Ciona intestinalis embryo, *Dev Growth Differ* ***50***, 521-529.
65. Neumuller, R. A., and Perrimon, N. (2011) Where gene discovery turns into systems biology: genome-scale RNAi screens in Drosophila, *Wiley Interdiscip Rev Syst Biol Med* ***3***, 471-478.
66. Lohmann, J. U., Endl, I., and Bosch, T. C. (1999) Silencing of developmental genes in Hydra, *Dev Biol* ***214***, 211-214.
67. Galliot, B., Miljkovic-Licina, M., Ghila, L., and Chera, S. (2007) RNAi gene silencing affects cell and developmental plasticity in hydra, *C R Biol* ***330***, 491-497.
68. Lawson, D., Arensburger, P., Atkinson, P., Besansky, N. J., Bruggner, R. V., Butler, R., Campbell, K. S., Christophides, G. K., Christley, S., Dialynas, E., Hammond, M., Hill, C. A., Konopinski, N., Lobo, N. F., MacCallum, R. M., Madey, G., Megy, K., Meyer, J., Redmond, S., Severson, D. W., Stinson, E. O., Topalis, P., Birney, E., Gelbart, W. M., Kafatos, F. C., Louis, C., and Collins, F. H. (2009) VectorBase: a data resource for invertebrate vector genomics, *Nucleic acids research* ***37***, D583-587.
69. Levashina, E. A., Moita, L. F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F. C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae, *Cell* ***104***, 709-718.
70. Eleftherianos, I., Millichap, P. J., ffrench-Constant, R. H., and Reynolds, S. E. (2006) RNAi suppression of recognition protein mediated immune responses in the tobacco hornworm Manduca sexta causes increased susceptibility to the insect pathogen Photorhabdus, *Dev Comp Immunol* ***30***, 1099-1107.

71. Werren, J. H., Loehlin, D. W., and Giebel, J. D. (2009) Larval RNAi in Nasonia (parasitoid wasp), *Cold Spring Harb Protoc* ***2009***, pdb prot5311.
72. Sullivan, J. C., Reitzel, A. M., and Finnerty, J. R. (2008) Upgrades to StellaBase facilitate medical and genetic studies on the starlet sea anemone, Nematostella vectensis, *Nucleic acids research* ***36***, D607-611.
73. Pankow, S., and Bamberger, C. (2007) The p53 tumor suppressor-like protein nvp63 mediates selective germ cell death in the sea anemone Nematostella vectensis, *PLoS One* ***2***, e782.
74. Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., and Patel, N. H. (2009) The crustacean Parhyale hawaiensis: a new model for arthropod development, *Cold Spring Harb Protoc* ***2009***, pdb emo114.
75. Zerlotini, A., Heiges, M., Wang, H., Moraes, R. L., Dominitini, A. J., Ruiz, J. C., Kissinger, J. C., and Oliveira, G. (2009) SchistoDB: a Schistosoma mansoni genome resource, *Nucleic acids research* ***37***, D579-582.
76. Bhardwaj, R., Krautz-Peterson, G., and Skelly, P. J. (2011) Using RNA interference in Schistosoma mansoni, *Methods Mol Biol* ***764***, 223-239.
77. Robb, S. M., Ross, E., and Sanchez Alvarado, A. (2008) SmedGD: the Schmidtea mediterranea genome database, *Nucleic acids research* ***36***, D599-606.
78. Newmark, P. A., Reddien, P. W., Cebria, F., and Sanchez Alvarado, A. (2003) Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians, *Proceedings of the National Academy of Sciences of the United States of America* ***100 Suppl 1***, 11861-11865.
79. Sanchez Alvarado, A., and Newmark, P. A. (1999) Double-stranded RNA specifically disrupts gene expression during planarian regeneration, *Proceedings of the National Academy of Sciences of the United States of America* ***96***, 5049-5054.
80. Lu, H. L., Vinson, S. B., and Pietrantonio, P. V. (2009) Oocyte membrane localization of vitellogenin receptor coincides with queen flying age, and receptor silencing by RNAi disrupts egg formation in fire ant virgin queens, *Febs J* ***276***, 3110-3123.
81. Cameron, R. A., Samanta, M., Yuan, A., He, D., and Davidson, E. (2009) SpBase: the sea urchin genome database and web site, *Nucleic acids research* ***37***, D750-754.
82. Song, J. L., and Wessel, G. M. (2007) Genes involved in the RNA interference pathway are differentially expressed during sea urchin development, *Dev Dyn* ***236***, 3180-3190.
83. Posnien, N., Schinko, J., Grossmann, D., Shippy, T. D., Konopova, B., and Bucher, G. (2009) RNAi in the red flour beetle (Tribolium), *Cold Spring Harb Protoc* ***2009***, pdb prot5256.
84. Ma, Y., Creanga, A., Lum, L., and Beachy, P. A. (2006) Prevalence of off-target effects in Drosophila RNA interference screens, *Nature* ***443***, 359-363.
85. Kulkarni, M. M., Booker, M., Silver, S. J., Friedman, A., Hong, P., Perrimon, N., and Mathey-Prevot, B. (2006) Evidence of off-target

effects associated with long dsRNAs in Drosophila melanogaster cell-based assays, *Nature methods* ***3***, 833-838.
86. Lin, X., Ruan, X., Anderson, M. G., McDowell, J. A., Kroeger, P. E., Fesik, S. W., and Shen, Y. (2005) siRNA-mediated off-target gene silencing triggered by a 7 nt complementation, *Nucleic acids research* ***33***, 4527-4535.
87. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BLAST+: architecture and applications, *BMC Bioinformatics* ***10***, 421.
88. Du, Q., Thonberg, H., Wang, J., Wahlestedt, C., and Liang, Z. (2005) A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites, *Nucleic acids research* ***33***, 1671-1677.
89. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate, *Embo J* ***20***, 6877-6888.
90. Anderson, E. M., Birmingham, A., Baskerville, S., Reynolds, A., Maksimova, E., Leake, D., Fedorov, Y., Karpilow, J., and Khvorova, A. (2008) Experimental validation of the importance of seed complement frequency to siRNA specificity, *RNA (New York, N.Y* ***14***, 853-861.
91. Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., Marshall, W. S., and Khvorova, A. (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets, *Nature methods* ***3***, 199-204.
92. Czech, B., and Hannon, G. J. (2011) Small RNA sorting: matchmaking for Argonautes, *Nature reviews. Genetics* ***12***, 19-31.
93. Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004) A protein sensor for siRNA asymmetry, *Science* ***306***, 1377-1380.
94. Elbashir, S. M., Harborth, J., Weber, K., and Tuschl, T. (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs, *Methods* ***26***, 199-213.
95. Chan, C. Y., Carmack, C. S., Long, D. D., Maliyekkel, A., Shao, Y., Roninson, I. B., and Ding, Y. (2009) A structural interpretation of the effect of GC-content on efficiency of RNA interference, *BMC Bioinformatics* ***10 Suppl 1***, S33.
96. Kirchner, R., Vogtherr, M., Limmer, S., and Sprinzl, M. (1998) Secondary structure dimorphism and interconversion between hairpin and duplex form of oligoribonucleotides, *Antisense Nucleic Acid Drug Dev* ***8***, 507-516.
97. Tafer, H., Ameres, S. L., Obernosterer, G., Gebeshuber, C. A., Schroeder, R., Martinez, J., and Hofacker, I. L. (2008) The impact of target site accessibility on the design of effective siRNAs, *Nature biotechnology* ***26***, 578-583.
98. Ender, C., and Meister, G. (2010) Argonaute proteins at a glance, *J Cell Sci* ***123***, 1819-1823.
99. Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J.

(2004) Argonaute2 is the catalytic engine of mammalian RNAi, *Science* ***305***, 1437-1441.
100. Aleman, L. M., Doench, J., and Sharp, P. A. (2007) Comparison of siRNA-induced off-target RNA and protein effects, *RNA (New York, N.Y* ***13***, 385-395.
101. Jackson, A. L., Bartz, S. R., Schelter, J., Kobayashi, S. V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P. S. (2003) Expression profiling reveals off-target gene regulation by RNAi, *Nature biotechnology* ***21***, 635-637.
102. Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L., and Linsley, P. S. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity, *RNA (New York, N.Y* ***12***, 1179-1187.
103. Petri, S., Dueck, A., Lehmann, G., Putz, N., Rudel, S., Kremmer, E., and Meister, G. (2011) Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects, *RNA (New York, N.Y* ***17***, 737-749.
104. Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P. D. (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway, *RNA (New York, N.Y* ***16***, 43-56.
105. Ameres, S. L., Hung, J. H., Xu, J., Weng, Z., and Zamore, P. D. (2011) Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two Drosophila Argonaute proteins, *RNA (New York, N.Y* ***17***, 54-63.
106. Vermeulen, A., Behlen, L., Reynolds, A., Wolfson, A., Marshall, W. S., Karpilow, J., and Khvorova, A. (2005) The contributions of dsRNA structure to Dicer specificity and efficiency, *RNA (New York, N.Y* ***11***, 674-682.
107. Vert, J. P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006) An accurate and interpretable model for siRNA efficacy prediction, *BMC Bioinformatics* ***7***, 520.
108. Fellmann, C., Zuber, J., McJunkin, K., Chang, K., Malone, C. D., Dickins, R. A., Xu, Q., Hengartner, M. O., Elledge, S. J., Hannon, G. J., and Lowe, S. W. (2011) Functional identification of optimized RNAi triggers using a massively parallel sensor assay, *Molecular cell* ***41***, 733-746.
109. Steinbrink, S., and Boutros, M. (2008) RNAi screening in cultured Drosophila cells, *Methods Mol Biol* ***420***, 139-153.
110. Raghavan, S., Williams, I., Aslam, H., Thomas, D., Szoor, B., Morgan, G., Gross, S., Turner, J., Fernandes, J., VijayRaghavan, K., and Alphey, L. (2000) Protein phosphatase 1beta is required for the maintenance of muscle attachments, *Curr Biol* ***10***, 269-272.
111. Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A., and Lewis, S. (2002) The generic genome browser: a building block for a model organism system database, *Genome research* ***12***, 1599-1610.
112. Yamaguchi, J., Mizoguchi, T., and Fujiwara, H. (2011) siRNAs Induce Efficient RNAi Response in Bombyx mori Embryos, *PLoS One* ***6***.

113. Bartscherer, K., Pelte, N., Ingelfinger, D., and Boutros, M. (2006) Secretion of Wnt ligands requires Evi, a conserved transmembrane protein, *Cell* ***125***, 523-533.
114. Bassik, M. C., Lebbink, R. J., Churchman, L. S., Ingolia, N. T., Patena, W., LeProust, E. M., Schuldiner, M., Weissman, J. S., and McManus, M. T. (2009) Rapid creation and quantitative monitoring of high coverage shRNA libraries, *Nature methods* ***6***, 443-445.
115. Li, L., Lin, X., Khvorova, A., Fesik, S. W., and Shen, Y. (2007) Defining the optimal parameters for hairpin-based knockdown constructs, *RNA (New York, N.Y* ***13***, 1765-1774.
116. Guberman, J. M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R., Chelala, C., Croft, D., Cros, A., Cutts, R. J., Di Genova, A., Forbes, S., Fujisawa, T., Gadaleta, E., Goodstein, D. M., Gundem, G., Haggarty, B., Haider, S., Hall, M., Harris, T., Haw, R., Hu, S., Hubbard, S., Hsu, J., Iyer, V., Jones, P., Katayama, T., Kinsella, R., Kong, L., Lawson, D., Liang, Y., Lopez-Bigas, N., Luo, J., Lush, M., Mason, J., Moreews, F., Ndegwa, N., Oakley, D., Perez-Llamas, C., Primig, M., Rivkin, E., Rosanoff, S., Shepherd, R., Simon, R., Skarnes, B., Smedley, D., Sperling, L., Spooner, W., Stevenson, P., Stone, K., Teague, J., Wang, J., Whitty, B., Wong, D. T., Wong-Erasmus, M., Yao, L., Youens-Clark, K., Yung, C., Zhang, J., and Kasprzyk, A. (2011) BioMart Central Portal: an open database network for the biological community, *Database (Oxford)* ***2011***, bar041.

Table 1 **Invertebrate model organisms amenable to RNAi**

Organisms	Database	Webpage	Database reference	RNAi reference
Acyrthosiphon pisum	AphidBase	http://www.aphidbase.com/aphidbase/	(*53*)	(*54, 55*)
Apis mellifera[1]	BeeBase	http://hymenopteragenome.org/beebase/	(*56*)	(*57, 58*)
Aplysia californica[2]	UCSC	http://genome.ucsc.edu/cgi-bin/hgGateway?org=Sea+hare	-	(*59*)
Bombyx mori	SilkDB	http://silkworm.genomics.org.cn/	(*60*)	(*61*)
Brugia malayi	Brugia Genome Project	http://ghedinlab.csb.pitt.edu/GhedinLab/Brugia%20malayi	-	(*62*)
Caenorhabditis elegans[3]	WormBase	http://www.wormbase.org/	(*50*)	(*4, 63*)
Ciona intestinalis[4]	JGI Genome Portal	http://genome.jgi-psf.org/Cioin2/	-	(*64*)
Ciona savignyi[5]	Ciona savignyi Database	http://www.broadinstitute.org/annotation/ciona/	-	(*64*)
Drosophila melanogaster[6]	FlyBase	http://flybase.org/	(*48*)	(*65*)
Hydra magnipapillata	Hydrazome	http://hydrazome.metazome.net	-	(*66, 67*)
Invertebrate vectors of human pathogens[7]	VectorBase	http://www.vectorbase.org/	(*68*)	(*69*)
Manduca sexta	Manduca Base	http://agripestbase.org/manduca/	-	(*70*)
Nasonia vitripennis	NasoniaBase	http://hymenopteragenome.org/nasonia/	(*56*)	(*71*)
Nematostella vectensis	StellaBase	http://nematostella.bu.edu/stellabase/	(*72*)	(*73*)
Parhyale hawaiensis	JGI Genome Portal	http://genome.jgi-psf.org/parha/	-	(*74*)
Schistosoma mansoni	SchistoDB	http://schistodb.net/schistodb20/	(*75*)	(*76*)
Schmidtea mediterranea	SmedGD	http://smedgd.neuro.utah.edu/	(*77*)	(*78, 79*)
Six ant species[8]	Ant Genomes Portal	http://hymenopteragenome.org/ant_genomes/	(*56*)	(*80*)
Strongylocentrotus purpuratus	SpBase	http://www.spbase.org/	(*81*)	(*82*)
Tribolium castaneum	BeetleBase	http://beetlebase.org/	(*49*)	(*83*)

[1] also including Bombus terrestris and Bombus impatiens
[2] Also available at BROAD http://www.broadinstitute.org/science/projects/mammals-models/vertebrates-invertebrates/aplysia/aplysia-genome-sequencing-project
[3] also including Caenorhabditis briggsae
[4] Also available via ENSEMBL
[5] Also available via ENSEMBL and http://mendel.stanford.edu/sidowlab/ciona.html
[6] also including 11 other Drosophila species (ananassae, erecta, grimshawi, mojavensis, persimilis, pseudoobscura, sechellia, simulans, virilis, willistoni, yakuba)
[7] Anopheles gambiae, Aedes aegypti, Ixodes scapularis, Culex quinquefasciatus, Pediculus humanus, Rhodnius prolixus, Glossina morsitans
[8] Atta cephalotes, Camponotus floridanus, Harpegnathos saltator, Linepithema humile, Pogonomyrmex barbatus, Solenopsis invicta

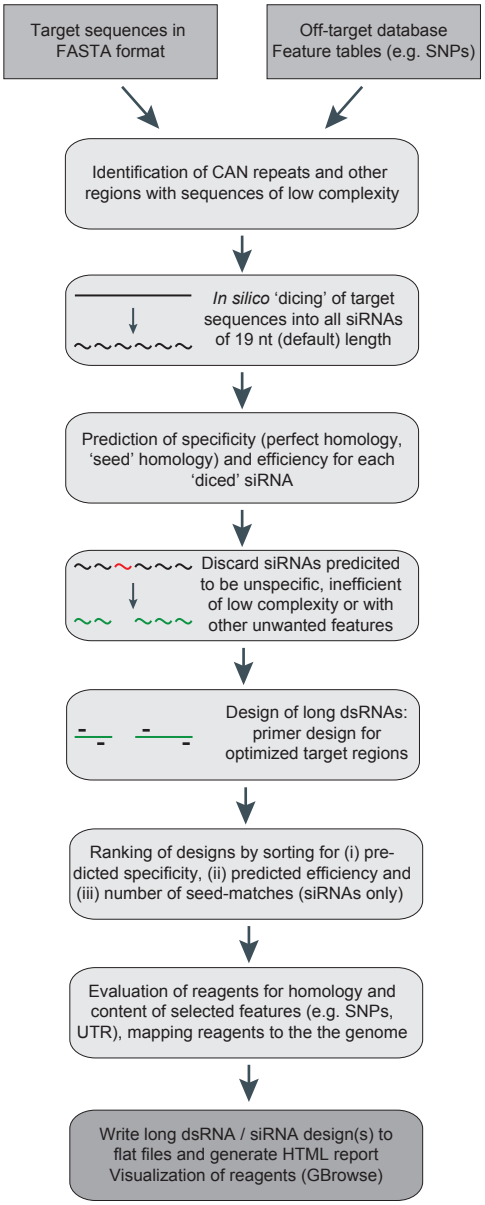


Figure 1
Horn and Boutros

Figure 2
Horn and Boutros

(a)

Exon flw:2

Design 1: flw:2_1

dsRNA information

Primer forward		Amplicon sequence
Sequence	taatacgactcactatagggATACGGATCGGGATCTGGTG	ATACGGATCGGGATCTGGTGCTGGTGCTCCTGGTGCCGGCGCTGGCCCAG GATCCAGCCATGATCCGCTCGACGATCTCACCGAACAGCAGCGCCGTCTG CTCCAGCAATACTCCATATCGCCGCCCAGCGAGACGATGATATCGGCCGA TGGCGACCAGATCACCGTCCATCATCCGCGCGAGGAGCCCAAGAAATCCA GATTGAAACTGCGCGATTTTT
Length [nt]	20	
Tm[°C]	62.034	
GC[%]	55.000	**Amplicon length [nt]**
Primer reverse		221
Sequence	taatacgactcactatagggAAAAATCGCGCAGTTTCAAT	**Amplicon location**
Length [nt]	20	X:10282633..10282853(+)
Tm[°C]	59.726	
GC[%]	35.000	

Primer pair penalty 2.3079

Target information

Intended target gene	FBgn0000711
Intended target transcripts (hits)	FBtr0071446 (203), FBtr0071447 (0)
Other targeted gene(s)	NA
Other targeted transcripts (hits)	NA

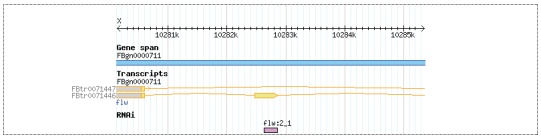
Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
203	203	0	0	26	46.57	0	0

Additional quality evaluation

Sequence homology (e-value)	FBgn0000711(1e-122)

Genome Browser



(c)

Exons flw:4-6

Design 1: X:10301000..10301520_1

dsRNA information

Primer forward		Amplicon sequence
Sequence	taatacgactcactatagggCTGTGAGCTTCACCTTCGGT	GTGTGAGCTTCACCTTCGGTGTGGATGTGGTCTCCAAGTTTTTGAACCGC CACGAGCTGGACTTGATCTGCCGTGCACATCAGGTGGGTGAACAAATTT AGTAATTTGACTGGATAATTCTAACGATATGTTGGCTTATTAGGTTGTGG AGGATGGCTATGAGTTCTTTGCGCGTCGGCAACTGGTCACGTTGTTCTCG GCGCCCAATTACTGTGGAGAG
Length [nt]	20	
Tm[°C]	60.307	
GC[%]	55.000	**Amplicon length [nt]**
Primer reverse		221
Sequence	taatacgactcactatagggCTCTCCACAGTAATTGGGCG	**Amplicon location**
Length [nt]	20	X:10301056..10301276(-)
Tm[°C]	60.647	
GC[%]	55.000	

Primer pair penalty 0.9541

Target information

Intended target gene	FBgn0000711
Intended target transcripts (hits)	FBtr0071446 (129), FBtr0071447 (129)
Other targeted gene(s)	NA
Other targeted transcripts (hits)	NA

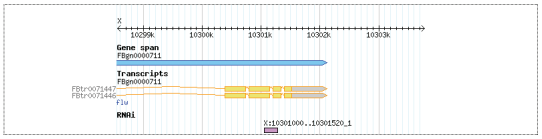
Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
203	129	0	74	32	51.20	0	0

Additional quality evaluation

Sequence homology (e-value)	FBgn0000711(1e-41)

Genome Browser



(b)

Exon flw:1

Target information

Intended target gene	FBgn0000711
Intended target transcripts (hits)	FBtr0071446 (196), FBtr0071447 (196)
Other targeted gene(s)	FBgn0085445; FBgn0086677
Other targeted transcripts (hits)	FBtr0112701 (2), FBtr0112699 (2), FBtr0112700 (2), FBtr0112694 (2), FBtr0112693 (2), FBtr0112702 (2), FBtr0112698 (2), FBtr0112696 (0), FBtr0112695 (0), FBtr0112697 (0); FBtr0088021 (1)

Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
196	193	3	0	44	54.25	0	0

Exon flw:3

Additional quality evaluation

Sequence homology (e-value)	FBgn0000711(2e-114);FBgn0004103(3e-21);FBgn0003134(6e-16);FBgn0003132(1e-10)

Exon flw:4

Target information

Intended target gene	FBgn0000711
Intended target transcripts (hits)	FBtr0071446 (207), FBtr0071447 (207)
Other targeted gene(s)	FBgn0004103; FBgn0003134
Other targeted transcripts (hits)	FBtr0082595 (3); FBtr0084642 (2)

Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
207	204	3	0	31	48.94	0	0

Exon flw:6

Target information

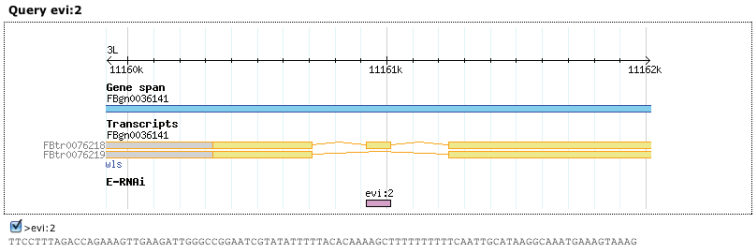
Intended target gene	FBgn0000711
Intended target transcripts (hits)	FBtr0071446 (200), FBtr0071447 (200)
Other targeted gene(s)	FBgn0033039
Other targeted transcripts (hits)	FBtr0086044 (1)

Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
200	199	1	0	50	58.12	0	0

Figure 3
Horn and Boutros

(a)



(b)

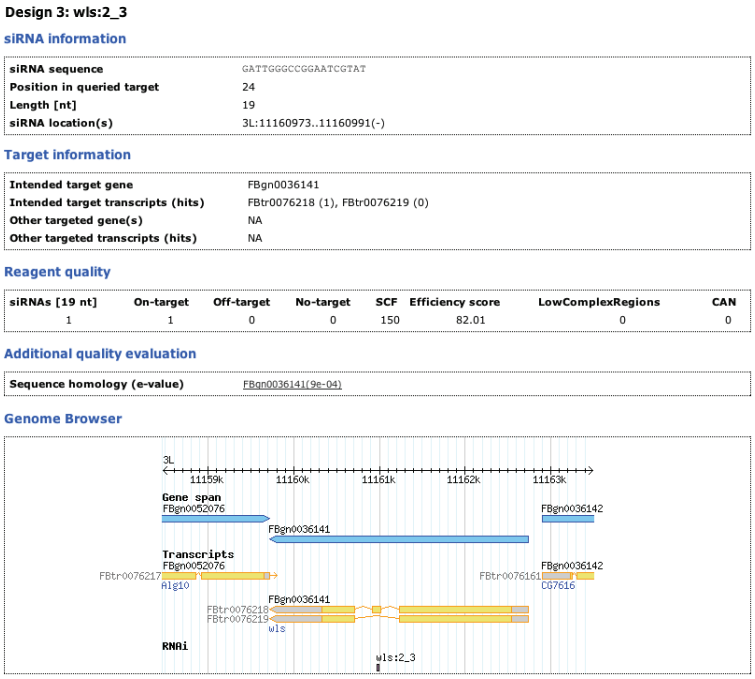


Figure 4
Horn and Boutros